DOCUMENT RESUME

ED 046 671 40 RE 003 292

AUTHOR Graubard, Paul S.; And Others
TITLE An Investigation Into the Use of Indigenous Grouping as the Reinforcing Agent in Teaching Maladjusted Boys to Read. Final Report.
INSTITUTION Yeshiva Univ., New York, N.Y. School of Education and Community Administration.
SPONS AGENCY Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
BUREAU NO BR-8-0174
PUB DATE Jun 70
GRANT OEG-8-08174-4353
NOTE 53p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS Behavioral Science Research, *Behavior Change, *Behavior Rating Scales, Emotionally Disturbed, Group Behavior, Maladjustment, Males, Peer Groups, *Peer Relationship, Positive Reinforcement, *Reading Achievement, Reading Instruction, *Reinforcement, Socially Deviant Behavior

ABSTRACT

An effort was made to assess the relative effectiveness of different reinforcement systems versus traditional teaching with maladjusted boys. The subjects were 60 boys in grades 6 and 7 who participated during language arts classes, but total data were gathered on 26 boys. The boys had histories of aggressive behavior and language problems. Treatment consisted of the systematic application of reinforcement principles, including the use of peer-group reinforcement. The subjects were observed for 10 seconds at the first minute of every 5 minutes using a behavior problem checklist designed for the study. Their reactions to group and individual reinforcements were noted and compared to reading achievement. Among the findings were that subjects improved consistently in appropriate social behavior and in most cases group reinforcement proved superior to individually distributed reinforcers. Significant correlations were obtained between characteristics measured by the checklist, academic gains, and social behavior. Tables and a bibliography are included. (MS)

ED046671

BEH BR-8-0174
PA 40
RE

FINAL REPORT
Project No. 8-0174
Grant No. OEG-8-08174-4353

AN INVESTIGATION INTO THE USE OF INDIGENOUS GROUPING AS THE REINFORCING AGENT IN TEACHING MALADJUSTED BOYS TO READ

Paul S. Graubard
Patricia Lanier
Hilde Weisert
Martin B. Miller

Yeshiva University
Ferkauf Graduate School of
Humanities and Social Sciences
55 Fifth Avenue
New York, New York 10003

U. S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

June 1970

Department of Health, Education, and Welfare

U. S. Office of Education
Bureau of Education for the Handicapped

RE003 292

ED046671

Final Report

Project No. 8-0174
Grant No. OEG-8-08174-4353

An Investigation Into The Use of Indigenous Grouping as The Reinforcing Agent in Teaching Maladjusted Boys to Read

Paul S. Graubard
Martin B. Miller
Patricia Lanier
Hilde Weisert

Yeshiva University
Ferkauf Graduate School of
Humanities and Social Sciences
New York, New York

June 1970

The research reported herein was performed pursuant to a grant with the Bureau of Education for the Handicapped, U.S. Office of Education, Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official position of the Bureau of Education for the Handicapped.

Department of Health, Education, and Welfare

U.S. Office of Education
Bureau of Education for the Handicapped

TABLE OF CONTENTS

Acknowledgements 2
Summary 3
Introduction 4
Method 5
Measures 7
Treatment 10
Results 12
Academic Achievement Effects 24
Discussions 34
References 39
Appendix A 42
Appendix B 44
Appendix C 47

Figures

1. Social Behavior for Classroom A over succeeding experimental phases and for each observational or treatment period..... 14
2. Social Behavior for Classroom B over succeeding experimental phases and for each observational or treatment period..... 16
3. Social Behavior for Classroom C over succeeding experimental phases and for each observational or treatment period..... 19
4. Social Behavior for Classroom D over succeeding experimental phases and for each observational or treatment period..... 23
5. Number of items attempted on Barnell-Loft by each of three classes over three experimental phases..... 29

Tables

1. Pre-and Post-Treatment Results on Spache Subtests, PPVT IQs for Participants in Study..... 25
2. Changes in Spache Diagnostic Reading Scales After Treatment..... 25
3. Intercorrelations of Gain Scores on Spache Subtests and Quay Checklist Score for Each of Four Behavior Categories..... 30
4. Intercorrelations Between Quay Checklist Category Scores and Number Attempted, Number Correct, and Percent Correct on Barnell-Loft Materials for Each of Three Designated Experimental Phases..... 31
5. Intercorrelation of Quay Categories and Social Behavior During Three Experimental Phases, for Each Period..... 33

Acknowledgements

The authors are grateful to the many people who made this study possible. The United States Office of Education was generous with its funds and co-operation. Dr. Herbert Goldstein, of the Department of Special Education at Yeshiva University, was also most helpful as were Drs. Montrose Wolf of the University of Kansas, Daniel O'Leary of the State University of New York at Stony Brook, Herbert C. Quay of Temple University, and Ann Duncan of Yeshiva University. Mr. Edmond James and Earl Miller of the New York City Board of Education were instrumental in seeing that this research was accomplished and they served as good friends as well as colleagues.

The teachers who participated in this study: Miss Flora Boyd, Mrs. Mary Brown, Miss Frieda Lichtenstein, and Mrs. Marie Perrault, were marvelous and they contributed so much to the study that it would have been impossible to do anything without them. We also thank the many children and teachers who gave so generously of their time and effort.

Beatrice Schwartz and Rose Friedman provided invaluable typing and clerical assistance. Al Piaget and Toni Petriconi, graduate students at Yeshiva University, among others, provided copious testing aid. Barbara Pentre was instrumental in helping us with practicalities of classroom application and training. Mr. Sidney Lipsyte, Director of Bureau of Maladjusted Children of the New York City Board of Education, once again proved himself an invaluable friend to those who do research. We also owe a debt to Dr. Samuel McLleland of the Bureau of Educational Review who gave approval for this study.

And finally, special thanks are due to our data collectors and assistants, Mrs. Viola Johnson, Mrs. Elsa Schaffer, and Mrs. Alice Richardson, and Miss Gretchin Mehegan, for without their diligent work, gregarious personalities and senses of humor, this project could not have succeeded.

Summary

Systematic application of reinforcement principles in a classroom setting, including use of a group of Ss' peers as a reinforcing agent, led to enhanced acquisition of reading skills, generally increased the incidence of desirable social behavior, and in some cases increased the rate of work. Ss were boys with a history of aggressive behavior and learning problems in four language arts classes of an intermediate school for delinquent and pre-delinquent youth. Ss improved consistently in appropriate social behavior. Initially striking differences between classes on academic rate vanished so that over a period of time, Ss in classes taught by new teachers were working as effectively as Ss in the classes of more experienced teachers. Social behavior and rate of work during non-reinforcement periods did not fall below baseline and in several classes held substantially above. In most cases group reinforcement proved superior to individually distributed reinforcers. Significant correlations were obtained between S characteristics as measured by the Behavior Problem Checklist, academic gain, and social behavior.

Introduction

Traditional approaches to teaching, or even maintaining in school, aggressive delinquent urban youngsters have so far neither insured these students minimal academic skills, nor stemmed the tide of dropping-out. Even where reduced class size (8 to 15 pupils) and auxiliary services are provided, there is no evidence of improved achievement or social behavior (Lipsyte, 1970). Conant (1961) holds that so long as education fails with this population, they are a sample of potential violence.

One source of this failure with maladjusted children has been found by several studies in the clash between the values of the school and the values of the delinquent group. If, as Cloward and Ohlin (1960) suggest, the school represents a value system unacceptable to urban delinquents, then the student's choice is between teacher approval and academic success, or the approval of his peers and membership in their group. When individuals in a residential treatment center hazarded the latter, the delinquent group successfully exerted its power in bringing them back into line with their mores (Polsky, 1962). Asch (1965) is one of many researchers who have demonstrated the potency of group norms in influencing opinions and actions. Evidently for the delinquent in this dilemma, the cost of success by school standards is often too high. Still, the traditional approach continues to be the "artichoke technique," with teachers trying to peel individuals away from the group over to the values of the school (Graubard, 1969).

Are there alternatives? Evidence from learning theory-based programs is promising. By rewarding him for academic achievement, Staats and Butterfield (1965) significantly raised the reading level of a 14-year-old delinquent. Wolf, Giles, and Hall (1968) substantially improved academic achievement in fifth and sixth graders, from an urban poverty area, by reinforcing their gains in an after-school remedial program with tokens exchangeable for trips, snacks, and games.

O'Leary, Becker, Evans, and Saudergas (1968) found sharp reductions in the disruptive behavior of seven children, in a second grade class of 24, under a token economy. By encouraging fewer aggressive statements, improved homework, and punctuality with tokens redeemable for privileges, Phillips (1968) significantly altered such behavior in three "pre-delinquent" boys in a residential treatment center. When Clark, Lachowicz, and Wolf (1968) paid five female dropouts for completing workbook assignments, achievement scores increased significantly. Group contingencies increased study behavior with two pre-schoolers (Bushell, Wrobel, and Midrallis, 1968) and eight delinquents (Graubard, 1969) in a laboratory environment.

The effectiveness of systematically applied reinforcement principles, particularly using the group as the reinforcing agent, has been established by these and similar studies. O'Leary (in press) has reviewed the literature on token economies and demonstrated its efficacy as a technique; certain questions, unanswered in his review, were explored in the current study.

The present investigators singled out several problems as in need of resolution before the promise of operant techniques in educating delinquent youth can be realized on more than a pilot basis in special schools and classes. This year-long investigation of operant techniques used with delinquent and pre-delinquent children addressed itself primarily to six questions:

1. Is teaching more effective with systematic reinforcement than without in a large-scale public school setting?

2. Can treatment-induced improvements in academic work rate and social behavior hold up over time, or are they temporary?

3. What happens to social behavior and academic work rate, in both reinforcement and non-reinforcement periods, when reinforcement is only given in selected periods and when it is removed altogether?

4. What kind of reinforcement -- that dispensed on a group, individual, or combination basis -- is most effective?

5. Can token economies be established in public schools for extended periods of time without using selected, trained teachers, and can the necessary skills be taught to regularly employed teachers as part of on-the-job training?

6. With what types of children can token economies be used most effectively?

Method

Subjects

The boys participating in this study had been placed in a special school following offenses such as assault, arson, and extortion, or release from State Training schools or hospitals.

Included in the analysis were only those Ss who remained in the experimental classes throughout the school year. Almost 60 children were worked with in this program, but the criterion of continuing in the experiment for the full year reduced the number included in the final analysis of social behavior to 26.

Two sixth and two seventh grade Language Arts classes, running for the first two 45-minute class periods of each school day, comprised the four Subject Classes. Reading skills, the presumed area of greatest deficiency, was the traditional curricular focus, with grammar, writing, class discussion, and outside reading added where student abilities allowed.

During this investigation, reading skills as taught through programmed materials formed the basis of the curriculum. Class groupings were determined by the clinical opinion of public school administrators and necessity, with no specific criteria outlined. Because of a highly mobile population and the school's service mandate (requiring adding a referred child or transferring out a child to regular school or residential placement) class composition fluctuated over the year. Registers for each of the four classes included from 8 to 15 students. With a high rate of truancy, average attendance was considerably below the number enrolled.

Teachers

Participation in the study was open to any interested language arts teacher, in line with the investigation's exploration of the effects of token reinforcement using average public school teachers in their normal setting.

The participating teachers had no previous training in applying reinforcement principles in the classroom, and little if any prior exposure to the approach. The investigators ran a series of after-school in-service workshops to train the teachers in token reinforcement methods; on the spot classroom observations, feedback, and modelling supplemented the afternoon sessions.

Each of the cooperating teachers held a New York City license. Experience varied: two from the group had worked with delinquent youth for at least two years; one had taught delinquent boys for one semester only; one, who took over one of the classes after more than a month of the term, was in her first formal teaching assignment.

Although the study began with five participating teachers and classes, one teacher was reluctant or unable to implement the reinforcement procedures outlined below, and his class was not included in the experiment.

Curriculum

The curriculum was held as nearly constant as possible during all phases (described below), with a set curriculum schedule followed during the daily Language Arts period. Basic materials were the programmed Barnell-Loft Specific Skills Series and the programmed SRA Reading Laboratories; each student was administered placement tests to determine his starting level in the material. The constant format - a reading selection followed by questions - and sequential presentation of the programmed materials allowed the most direct means to chart ongoing changes in reading skills, work output, and accuracy.

Measures

Two classes of behavior were included as dependent variables in this investigation. One was social behavior, observed systematically via a coded checklist of "study" (appropriate) and "deviant" (inappropriate) categories of activity, as developed by Becker, Madsen, Arnold, and Thomas (1968).

Procedure

Each S was observed serially for 10 seconds at the first minute of every five minutes, following the Becker, et al, method by a trained observer seated at the back of the classroom. At least four observations per period were obtained on each child during individualized seatwork. Because the Becker coding categories were developed for use with younger children, several additions were made to encompass the deviant possibilities of the delinquent adolescent. These additions, however, did not affect the ratio of deviant to study behavior, since all coded observations were subsumed into the larger categories of appropriate study behavior and inappropriate deviant behavior, yielding percentage figures for average frequency of individual S's study behavior. Inter-observer reliability, calculated as total number of cells times 100, was checked weekly during baseline, and throughout the year at varying intervals. Agreement never fell below 82%, with the year's average at 90%.

Reading achievement, as indicated in several measures, constituted the second class of dependent variables studied. Measures included: 1) Reading gain scores on the Spache Diagnostic Reading Scales (1963), an individualized measure with indices of word recognition, listening comprehension, and oral and silent reading; and 2) Rate of frames completed in the Barnell-Loft Specific Skills Series (Boning, 1965).

Behavioral Checklist

Another aspect of this project was to gather prediction data about what kind of youngster could profit the most from token reinforcement.

One basis for prediction is the grouping of children into given categories, and consideration of criterion differences in terms of such categorization. However, grouping children presents a problem, the most serious being the typical lack of reliability of many classification schemas (Eiduson, et al, 1966).

Fortunately, Quay (1966c, 1966b, and Quay, Morse, and Cutler, 1966) has described an objective method for classification of children with behavioral disorders. The method initially developed by Peterson (1961) has been used in previous classroom research (Graubard, 1968) and appears to be relevant to the problem of classifying children within the public school framework.

Briefly, the classification schema involves a behavior checklist, with all behaviors practically observable in a child's case history folder or similar resources. The final Behavior Problem Checklist as ultimately derived by Peterson, et al., consists of 58 items.

The Behavior Problem Checklist (see Appendix B) categorizes observable behaviors and requires that the judge or rater sees the child in living situations or takes information from case histories. Thus, inferential attributes are minimal. Several studies (Quay, 1966a, Quay, Morse, and Cutler, 1966) have shown that three independent dimensions account for about two-thirds of the variance of the interrelationships among the problem behaviors.

The four factors extracted in previous (Quay, 1966b) research reveal that the first dimension of the scale uncovers (1) aggressive, hostile behavior, and is usually labelled "conduct disorder," a form roughly analagous to "unsocialized aggression," or "psychopathy." The second dimension represents anxious, depressed, introvertive behavior, and can be labelled "personality problem" (P). The third dimension involves disinterest, apathy, daydreaming, and passivity. The labels of "inadequacy" (I), and "immaturity," have been used to describe this dimension (Quay, Morse, and Cutler, 1966a). Quay(1963) has suggested that a fourth dimension -- socialized delinquency (SD)-- applies to a proportion of inner city youth who are not disturbed in the classical sense, but who are at odds with middle-class schools and teachers.

In order to utilize this scale to classify the Ss, each teacher, after one month of classroom contact with pupils, rated them on the Behavior Problem Checklist. Ss were scored as to whether they displayed a problem or didn't ; degree was not taken into account.

Design

The study originally conceived to assess relative effectiveness of different reinforcement systems, and of these versus traditional teaching, compared across classes, through an ABA operant design. This design, with each S serving as his own control, using a baseline period of traditional teaching (A_1), application of contingencies (B), and return to baseline conditions (A_2), was altered because of two factors:

1. Through baseline data the classes quickly emerged as differing widely in frequency of deviant S behavior, and in categories of teacher behavior, so that subsequent analyses ignoring these differences would be of questionable value. Also, Ss' return to baseline level behavior late in the term—i.e., after treatment - seemed unlikely, since many of the changes produced during treatment would be irreversible. When a student has acquired a new reading skill, expanded his concentration span, or begun to enjoy the intrinsic rewards of improved academic achievement, he is no longer the same child. "Valuable behavior, once set up, may no longer be dependent upon the experimental technique which created it" (Baer, Wolf, & Risley, 1968).

2. Class fluctuation in a service-mandated public school, as mentioned above, created shifting class dynamics with each addition or transfer out of a child. To reduce the effect of irreversible changes and of changing class composition across time, a "multiple baseline" technique was incorporated allowing each class to be compared with itself across periods. This technique is fully described by Wolf, Giles, and Hall (1968).

Following this design, baselines were established on target (i.e., to be modified) behaviors against which changes could be evaluated. Two baselines for each of the two class periods (9 - 9:45; 9;45 - 10:30) were obtained and then the experimental variable was applied to the target behaviors during one of the periods and not during the other (Phase II, Condition B, 9 - 9:45; Condition A, 9:45 - 10:30). If the experimental variable was effective, a change would be produced in target behaviors in the class period where it was applied, and little or no change would be produced in the period of continuing baseline. Then, treatment was removed from the first period and applied to target behaviors in the second. If target behaviors during second period changed at that point and those during first period shifted in the direction of their baseline level, evidence would be mounting that the experimental variable was effective, and that the prior

change was not simply a matter of coincidence. (Wolf, personal communication, 1968).

Treatment

Phase I, Baseline. To evaluate later changes, a four week baseline period was instituted after a three week period of habituation. During baseline teachers followed their traditional teaching methods (e.g., teacher praises, exhortation, tests, marks, punishment, etc.).

Phase II, Experimental. During Phase II a token economy was established in each class for at least part of the two language arts periods. Students contracted with teachers for rewards (back-up reinforcers); the reinforcers were made contingent upon following posted rules of social behavior. The four rules, deemed necessary by teachers for study to occur, were: sitting in seat, raising hand for permission to speak, paying attention to task, and completing individually assigned programmed materials with a specified degree of accuracy, within specified time limits.

Several variations of a token system were instituted. In some cases, i.e., Group + Individual + Group Reinforcement (G+I+G), behavior points were earned only if the entire group followed the rules, individual points were earned for academic work, and every one had to earn a minimum number of points before the group could cash in its tokens. In the Group + Individual + Individual Reinforcement condition (G+I+I), behavior points were dependent on the performance of the group. But the contract specified that when an individual reached the requisite points for his choice, he would receive it immediately, and begin working towards his next chosen reward.

During the individual reinforcement situation (I), work points were earned by individual Ss according solely to completion of a requisite number of frames in the programmed material. There were also individually-given behavior points: whoever followed the rules earned points regardless of his classmates' behavior.

Procedure during reinforcement

Rewards were earned through the accumulation of points. Points were entered daily in individual bankbooks and posted on blackboard tallies, with a specified number of points necessary to "buy" the chosen back-up reinforcer. Work points were earned on the basis of number of problems correct, i.e., rate, as well as percent accurate; points earned for social behavior were dispensed by the teacher's

ringing a short-ring timer set by her for varying intervals during the period. In the group-contingency classes, when the timer sounded and all Ss were following the behavioral rules, they were reinforced by all Ss earning two points. When the clock rang and one or more Ss was not following the rules, no one earned points. The teachers were instructed to: (1) repeat the rules, (2) frequently give oral feedback on points earned throughout the period, with final posting specifically labelled (from tally kept on wrist counter), (3) give children bankbooks to enter tally at the end of the period, and (4) keep a daily record of points earned and work done.

Unless it threatened safety, all deviant behavior was to be ignored. In dangerous instances, teachers were to follow regular school procedures, which usually meant sending children to the principal.

Another experimental phase to the project was implemented to control and examine the possible effects of time and time-sequence, (e.g., do Ss do better first thing in the morning regardless of teaching conditions?). In Part 2 of Phase III, the token economy system was switched to the second period of the day, and traditional teaching now occupied the first period. Each class continued with the condition with which it had begun. Curriculum was unchanged, although reinforcement was not contingent on social behavior and academic performance on SRA (1963), rather than Barnell-Loft. The curricular materials were found to be interchangeable regarding rate of completed work and social behavior emitted.

Non-Contingent Reinforcement (NCR) Phase

Between the fourth and fifth months of the study a phase of NCR was introduced in all classes. This phase was implemented to explore the effects on academic performance and social behavior of discontinuing contingent reinforcement. If contingent reinforcement had over five months effected academic and behavioral gains, would its withdrawal return Ss' performance to baseline level or below?

During NCR each child was awarded, at the beginning of the day, the average number of points earned daily during the previous phase. To make it clear to Ss that points earned were no longer contingent on work or following classroom rules, these points were attached to an academically irrelevant but concrete behavior, i.e., two classes were to appear in class with matching socks, while the other two classes were to appear with clean hands (a teacher preference). Teachers were instructed to state clearly to Ss that during this phase this and this only was the means by which they could earn points, that though Ss were expected to continue

working hard and behaving, they would receive no points for this and points had to be given out before the period started. Teachers were advised to use whatever techniques -- short of contingent token reinforcement -- they found effective in encouraging study behavior. More particularly, they were encouraged to return to their original method of classroom management and teaching as much as possible.

RESULTS

Initially, data were cast into each of two basic designs which permitted simultaneous consideration of all participating classes, irrespective of special contingency arrangements for any given class. The first of these analyses examined weekly averages for social behavior over the first 17 weeks of the program, with two data points per week, one the average for Period 1 (9:00 - 9:45), and the other for Period 2 (9:45 - 10:30), for each of four classes. There were thus three dimensions: weeks, periods, and classes. These data were subjected to a "mixed" analysis of variance (Lindquist, 1953, pp. 292-297), with weeks and periods treated as "within" effects, and classes as a "between" effect.

The results of the foregoing analysis showed a highly significant ($p < .001$) weeks x periods x classes interaction ($F=9.35$, with 48 and 384 df), as well as several significant lower-order interactions and main effects. Subsequent breakdown of the design and component analyses revealed that there were significant differences between inconsistencies appearing even during the baseline period. For example, while one class showed considerable stability of social behavior during baseline, another class was highly erratic and irregular during these early weeks of the project, i.e., prior to the introduction of contingency arrangements.

A second general analysis considered differences in social behavior across classes with data summarized into three phases: (I) average social behavior over the first four or "baseline" weeks, (treatment) over the three weeks prior to non-contingent reinforcement (NCR) arrangements, and over three weeks of NCR. Once again, two data points were represented for each phase, i.e., one for each of the two periods. Since one class did not participate in NCR for reasons explained elsewhere, data for only three classes were considered in this analysis, which also was treated as a Lindquist mixed (Type VI) ANOVA, (phases and periods as "within" effects, classes as a "between" effect.)

The results of this phase x period x class analysis showed an interaction of periods and classes that was significant beyond the

.001 level (F=15.15, 2 and 17 df). All other interactions were not significant, while main effects for phases and for classes were highly reliable ($p < .001$). In view of the significant period x class interaction, a breakdown of the larger design to appropriate component sub-designs revealed that one class showed overall significant differences between periods 1 and 2, i.e., with data collapsed across phases, while differences between periods for the other two classes were not significant. Analysis of the main effect for phases showed that, collapsing across classes and periods, social behavior for the baseline phases was significantly lower than the phase during the three weeks prior to NCR, but missed significance at the .05 level when compared with the NCR Phases. The main effect for classes showed one class to be significantly higher in general than the other two.

These analyses generally underscore the distinctive qualities of each participating class in the project, and these in turn are undoubtedly reflective of idiosyncratic qualities in both teachers and children which are hardly surprising in studies of this nature. Therefore, the remainder of our analyses were directed toward consideration of each class in an attempt to uncover particular effects within them, especially in view of variations in contingency arrangements that were specifically applied to each class.

These class-specific analyses considered all phases of the project for any given class. Also taken into account were the daily periods. Thus, in each such analysis, data were cast into an AxBxS design (Lindquist, 1953), i.e., phases by periods by Ss. In this context, the term "phase" refers to any variation of reinforcement conditions for one or more weeks, including the switching of a given reinforcement condition from one period to another along with the removal of reinforcement for a period, thus creating a reversal effect, or the contrasting of two different reinforcement arrangements over several weeks (e.g., one arrangement for each period.)

The first of the class-specific analyses is with reference to Classroom A. Average social behavior effects for each phase for Classroom A are presented in Figure 1. Each phase is indicated on Figure 1 and in succeeding figures by a Roman numeral.

> Wherever "simple"effects were examined, such as differences between phases or differences between periods within phases, the appropriate error term derived from the full table was used. In this regard, per Lindquist's recommendation, critical differences (d) were generated for each such simple comparison. Figure 1 and succeeding figures indicate the d necessary for significance at the .05 and .01 levels which apply to the data plotted on the figures.

Figure 1.

Social behavior for Classroom A over
succeeding experimental phases
and
for each observational or treatment period.

figure 1

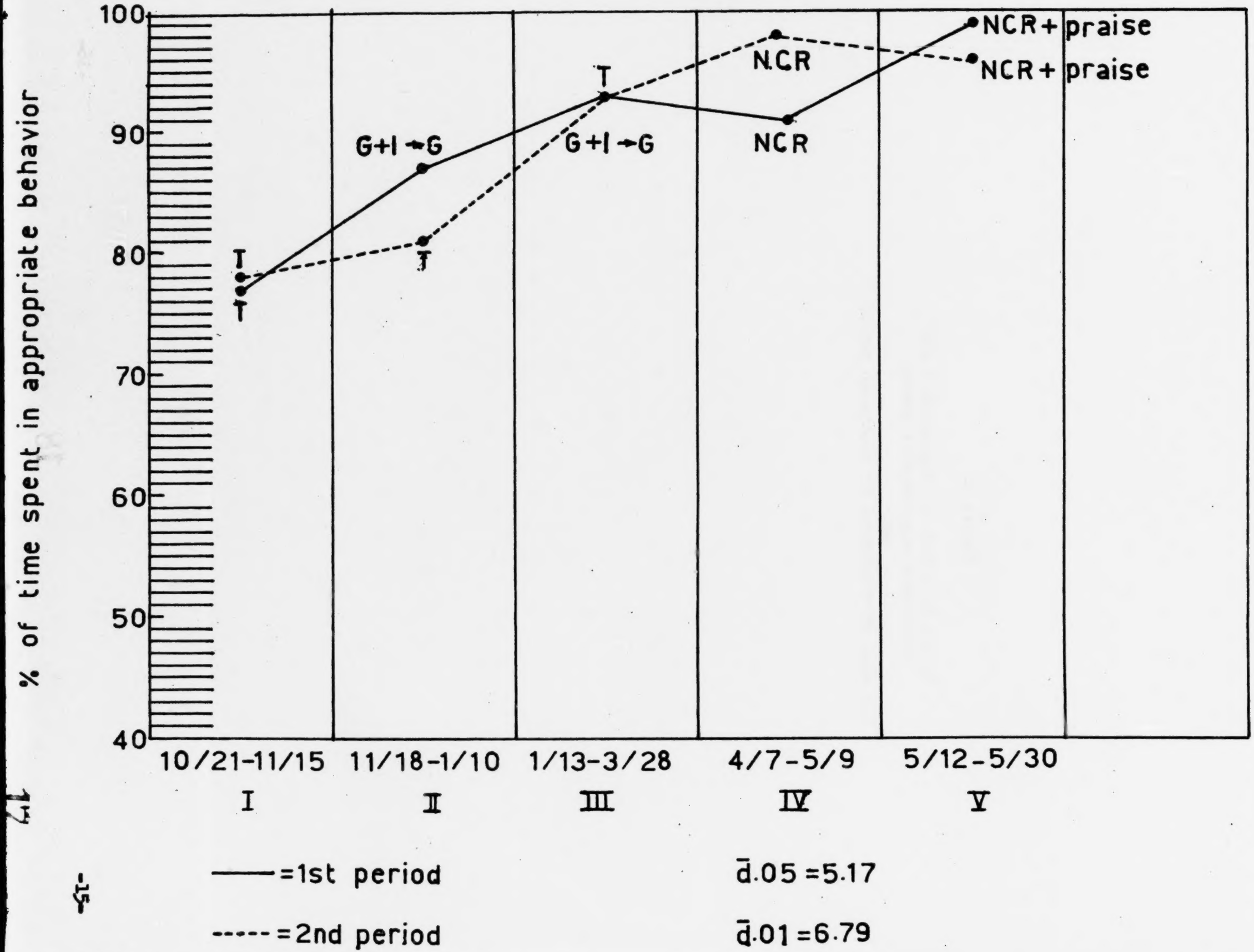

100
90
80
70
60
50
40
% of time spent in appropriate behavior
T
T
T
G+I→G
T
T
G+I→G
NCR
NCR
NCR+ praise
NCR+ praise
10/21-11/15
I
11/18-1/10
II
1/13-3/28
III
4/7-5/9
IV
5/12-5/30
V
——=1st period
-----=2nd period
d̄.05 =5.17
d̄.01 =6.79

Figure 2.

Social behavior for Classroom B over
succeeding experimental phases
and
for each observational or treatment period.

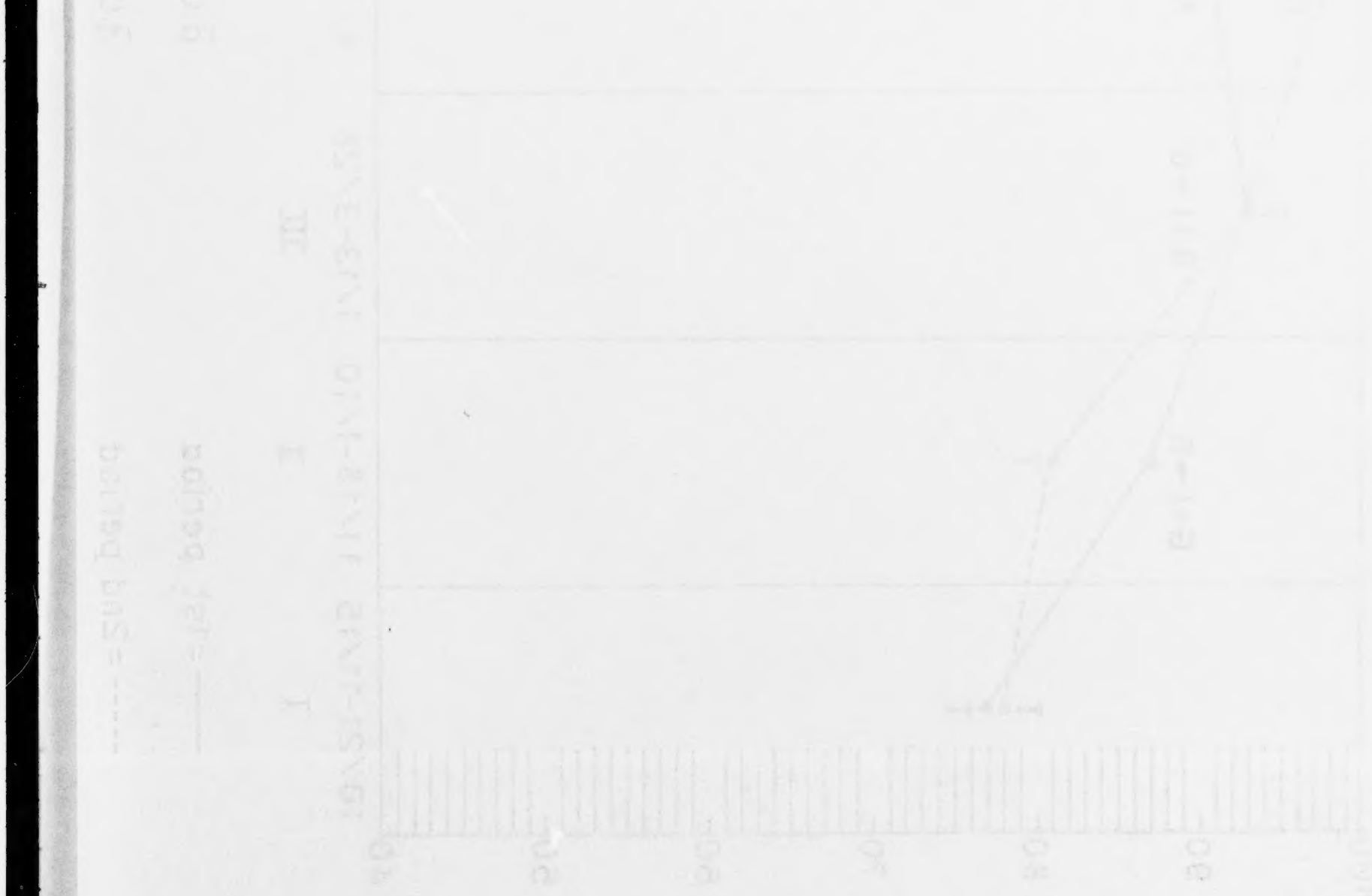
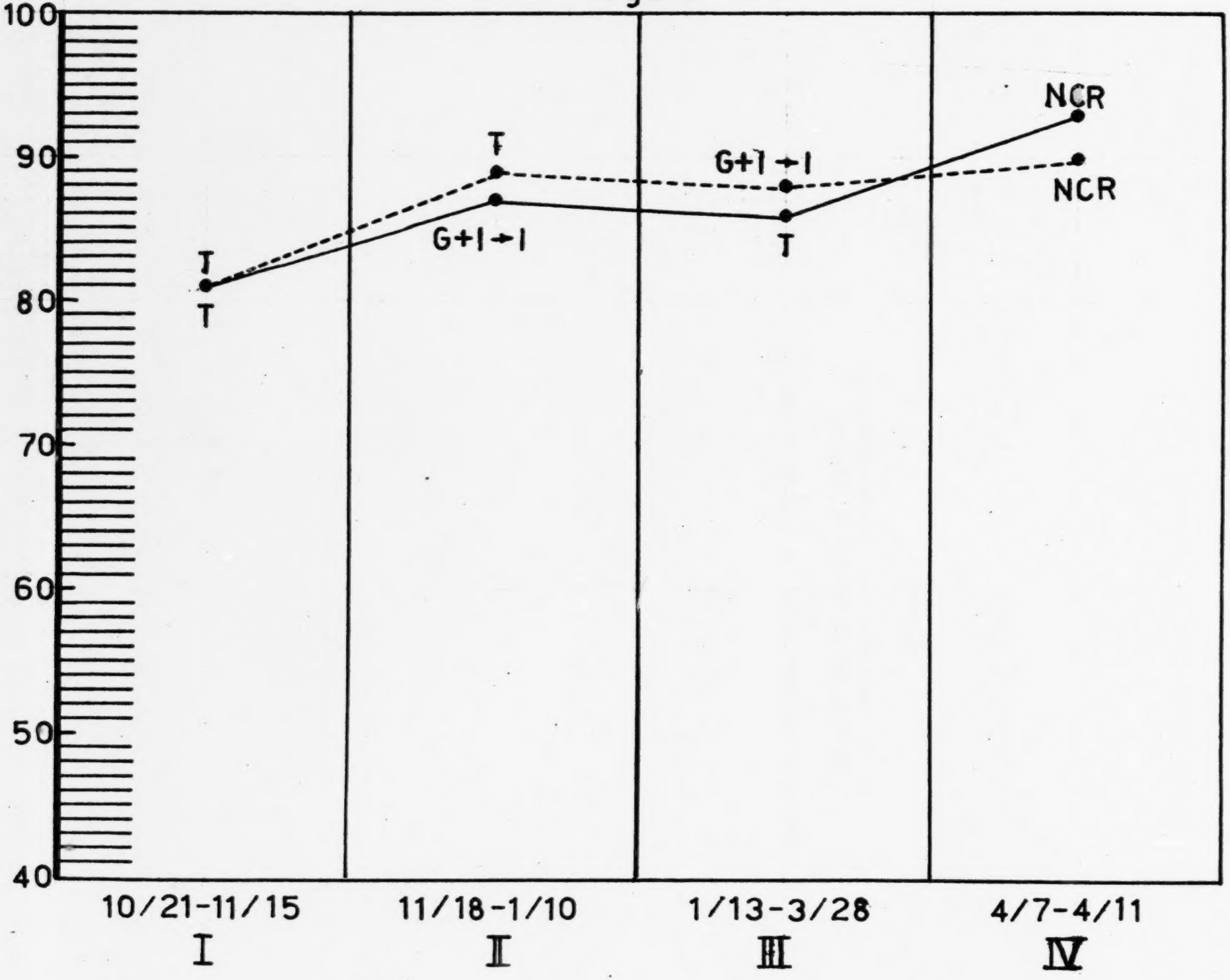


figure 2



—— = 1st period

----- = 2nd period

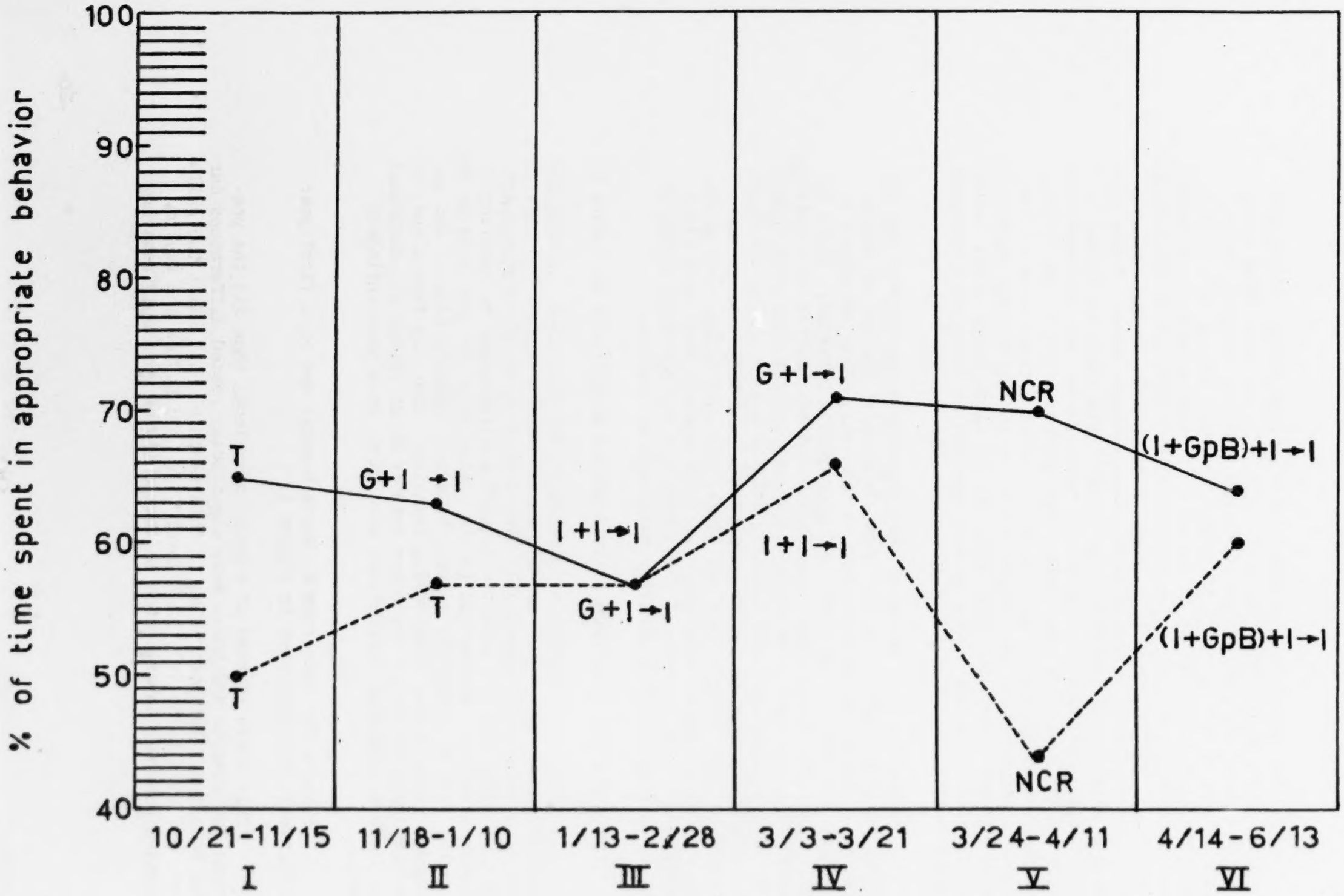
$\bar{d}_{.05} = 10.84$

19

Figure 3.

Social behavior for Classroom C over
succeeding experimental phases
and
for each observational or treatment period.

figure 3



——=1st period $\bar{d}_{.01}=15.41$

------=2nd period $\bar{d}_{.05}=11.13$

In this class, which started out on high level, there were initially no significant differences across periods. When group reinforcement was introduced in Period 1 during Phase I, there was a significant increase in social behavior during this period, but it did not generalize to Period 2 so that on both a horizontal and vertical level, group reinforcement was superior to traditional teaching. During Phase III, group reinforcement during Period 2 proved superior to Period 2 during the previous Phase. There was also a significant increase under traditional teaching during Phase III. Although observation of teachers was not systematic in this study, it is our impression that gains accrued to the traditional teaching period, because this teacher avoided some of the direct confrontation techniques which she had previously employed. When NCR was introduced this class showed significant gains during Period 2 so that the period proved superior to Group Reinforcement prior to the introduction of NCR and to NCR during Period 1.

During the last phase of this experiment the teacher was instructed to add praise to her usual teaching. The use of praise seemed to result in an increment so that this class continued to work at the level of at least 96% appropriate behavior. Thus, in this class the token system possibly aided the teacher in acquiring more effective ways of handling behavior, and teaching with tokens was more efficacious than teaching without tokens during the first semester. During the latter part of the year this teacher continued to gain control over her class, and when tokens were given on a non-contingent basis there was more control over this class than even during baseline. Thus, behavior did not deteriorate in this class and in fact behavior continued to improve.

Results for Classroom B are depicted graphically in Figure 2.

In this class, group reinforcement was also used. During the baseline phase there were no differences across periods and behavior was at a relatively high level. The introduction of reinforcement in this class did not effect a significant increment on behavior during baseline. During NCR (which lasted only one week because the teacher went on maternity leave), behavior remained high -- and was significantly higher than during baseline. Thus, in this class, behavior improved during the token reinforcement system and continued to improve even when reinforcers were given on a non-contingent basis.

Results for Classroom C, whose teacher was on a first-year assignment, are presented in Figure 3.

This class started at a much lower level than did the previous two classes and there were significant period differences during baseline. It appeared as if this teacher could start the children off at a relatively high level (for her) but as the day progressed behavior deteriorated so that children were only behaving

appropriately 50% of the time. When Group Reinforcement was introduced not very much happened because of the inconsistent way that it was applied, and there were no differences in these classrooms between group and individual reinforcement or between reinforcement conditions and non-reinforcement until Phase IV. It seemed to take the teachers about five months to learn not to give "second chances" or not to add "buts" to praise.

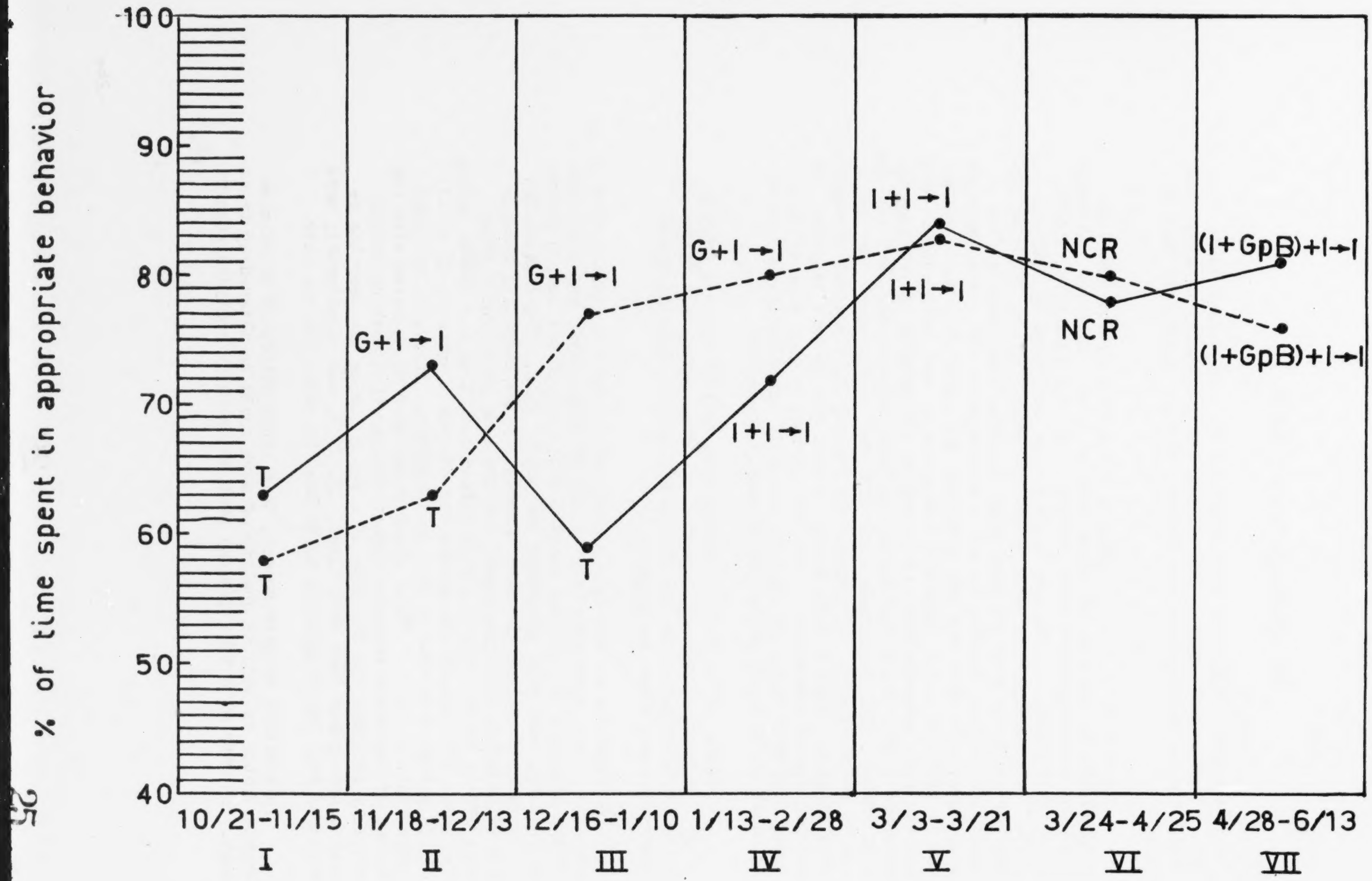
There seems to have been almost a cumulative effect and when reinforcement techniques began working appropriate behavior increased to 71% during group reinforcement and 65% during Period 2-individual reinforcement. This is the first time that this class significantly improved over baseline. During NCR Period 2 a reversal occurs and there is a significant decrement in behavior when compared to contingency teaching, and while behavior during this period is lower than baseline, it is not reliably lower. In addition when tokens were reinstated the behavior showed a significant improvement. Interestingly enough, during NCR the significant difference between periods re-emerges and the teacher and/or class cannot sustain the behavior of Period I. When reinforcement was re-introduced the differences between periods disappeared and during Period 2 reinforcement again proved superior to NCR teaching. Thus, an analysis of this class showed that a token system can support a teacher over time so that the token system can help to maintain the teacher's own "optimal" level of functioning. It appears that with this teacher the token system more than coincidentally correlated with appropriate behavior. It seemed to have actually produced these changes and was able to maintain this level for almost a full academic year. We also see in Class C during Period I that the G condition was higher than the I condition for Phases II and IV, while during Phase III (I condition), there was a decrement in behavior.

Figure 4 shows Classroom D, the class of a teacher with just one year of experience with regular class pupils. The teacher in Classroom D also starts out with social behavior at a relatively low level, although more consistent than Classroom C as no period differences emerges during baseline. When group reinforcement is introduced it proves significantly higher than traditional teaching, both on a horizontal and vertical level, and that this is not a function of time of day is shown by the reversal during Phase III of this experiment. When group vs. individual reinforcements are compared (from 1/13 to 2/28) group reinforcers emerge as significantly more powerful than individual reinforcers. When NCR is introduced the direction appears to go down, but not significantly, and NCR remains significantly higher than baseline conditions. It appears then with this class that tokens are clearly effective in changing pupil behavior, and the teachers learned to use the successful techniques in NCR in contrast to Classroom C where there was a decrement in behaving during NCR. In addition, in this class as in all classes behavior did not deteriorate over time or during phases when token reinforcement was not used.

Figure 4.

Social behavior for Classroom D over succeeding experimental phases and for each observational or treatment period.

figure 4



——=1st period

----=2nd period

differences between phases:

$\bar{d}.05 = 7.40$

$\bar{d}.01 = 9.98$

differences for periods within phases:

$\bar{d}.$=8.61 @ .05

$\bar{d}.$=11.60 @

Academic Achievement Effects

Academic gains were made apparent by achievement tests. Table 1 shows the summary data of the initial Spache testing, the final Spache testing, and the Peabody Picture Vocabulary Test IQ scores.

Inspection of Table 1 reveals the generally low level of functioning in reading for these special school children and these findings are consistent with previous research (Graubard, 1968). Also interesting are the low IQ scores as indicated by PPVT results. This measure is reported in the Manual (Dunn, 1959) as being particularly good for assessing children with reading disabilities because reading is not a component of the actual testing procedure. The test was not developed for inner city adolescents and is probably not a sensitive indicator of what urban adolescents can achieve; nevertheless, it is fairly independent of reading, as Graubard (1967) found Pearson correlations between the PPVT and the WISC Verbal, WISC Performance and WISC Full Scale were .59; .24; and 56 respectively. Of course, the Peabody is a measure of verbal intelligence. What is noteworthy is that the Ss were labelled and had manifested behavioral disorders; perhaps a good many of them could have been labelled educable mentally retarded instead of or in addition to being called behaviorally disordered.

A number of comparison groups were used to measure changes of Ss. Primarily, each S served as his own control and difference scores were computed for the Spache Recognition Reading Scale. These data are shown in Table 2.

Elaboration of the data summarized in Table 2 reveals that 18 subjects, i.e., more than half of the showing sample but about 38% of the children the project started with, exceed one year's growth in reading, with most exceeding two years' gain. These gains may be considered against the data compiled by New York City Schools for Maladjusted Children report for the same year, which shows that 67% of Bureau school pupils gained less than six months during the year, 25% between six months and one year, and only 8% of all pupils gained more than a year in 10 months. (Lipsyte, personal communication, 1970.) These figures can also be compared with the Bureau of Education Research report showing that reading scores actually declined for the city as a whole during the year the research took place (New York Times, 2/15/70), and considerably more children fell below national norms than is usual for the city.

In addition to gain scores, this study attempted to examine whether rates as well as quality of work could be altered by treatment. In line with this general problem, several analyses were run

Table 1

Pre-and Post-Treatment Results on Spache Subtests,
and
PPVT IQs, for Participants in Study (N=24)

	Spache-Pre-Test			Spache Post-Test			PPVT
	Wd.Rec.	Oral Read.	Silent Read.	Wd.Rec.	Oral Read.	Silent Read.	IQ
M	3.99	4.30	4.71	5.27	6.09	6.86	81.69
SD^2	6.08	7.09	8.69	5.75	9.12	7.68	35.92

Table 2

Changes in Spache Diagnostic Reading Scales After Treatment (N=22)

	Spache Scale		
	Word Recognition	Instructional Level (Oral)	Independent Level (Silent)
ΣX	33.1	37.7	72.8
ΣX^2	58.91	100.83	302.88
Mean gain	1.50 yrs.	1.71 yrs.	1.46 yrs.
t	10.20**	6.13**	7.11**

** $p < .01$

with respect to each of the three dependent variables culled from the Barnell-Loft materials: (1) number of items attempted, (2) number of items correct, and (3) percent of items correct. Each of these types of data were analyzed in two general ANOVA designs, one for the first 17 weeks as one dimension and classes as the other dimension, while in the second general design, three "phases" (baseline; treatment; NCR) were treated as one dimension and classes was the second dimension. These analyses were similar to those conducted with respect to social behavior, except that no distinctive data were available for Periods 1 and 2 in Barnell-Loft analyses.

The analyses for "number attempted", which of course is analogous to a rate analysis, showed significant class by weeks (in the 17-week case) and class by phase interactions. It is hardly surprising that there should be differential fluctuations in rate as a function of classroom, as in the 17-week analysis, but the results of three-phase analysis took a form that had not been altogether expected. In effect, one class (A) actually showed a significant rate decrement from baseline to treatment phases. Further reflection leads us to the following. Initially, one might expect that because of experience, there would be significant differences between classes in rate of work attempted on Barnell-Loft materials precisely as found. The differences between Class A and C were significant at the .01 level. It will be recalled that there were striking differences between classes in level of social behavior as well. During treatment significant differences are again found between Classes A and D but these differences are probably an artifact of a restriction on the amount of programmed materials that Ss in Class A were allowed to complete. Such restrictions were not imposed on Classes C and D and their Ss could complete as much work as time allowed. This restriction on Class A was removed during NCR and there were no differences between any of the classes during this phase. Thus, a plausible interpretation is that the treatment was instrumental in washing out differences between experienced and inexperienced teachers.

Figure 5 presents graphic data for the number of items attempted on Barnell-Loft material over the various treatment phases.

Analyses of "number correct" yield the same types of effects as obtained for rate, but this should be expected since opportunity to be correct is a direct function of rate of attempts. "Percent correct" analyses showed a significant weeks by class interaction in the 17-week analysis, but no significant effects in the three-phase analysis. Further examination of the 17-week data for "percent correct" reveals that no class shows any significant effect for this variable over the 17 weeks of data analyzed. The significant interaction is explained by some fluctuations in spread between classes for different weeks, but nothing of systematic consequence. Thus, it is apparent that our treatments do not have systematic effects on the general quality of academic work performed, at least as defined in terms of percent correct on the Barnell-Loft, while rate attempted and therefore absolute rate correct may be affected by the token system. This latter point seems especially true when certain individual difference factors are taken into account.

Because there are no reliable differences between NCR and Baseline phases, this data also supports the contention that Ss will not slack off efforts from baseline levels when contingent reinforcement is removed.

Groupings as Predictors

Correlations were run between the behavior dimensions of the Quay Behavior Problem Checklist and several academic and social measures.

The following table (Table 3) shows the correlations between gain scores on the Spache Diagnostic Reading Scales and behavior categories as derived from the checklist.

Figure 5.

Number of items attempted on Barnell-Loft by each of Three classes over Three experimental phases.

figure 5

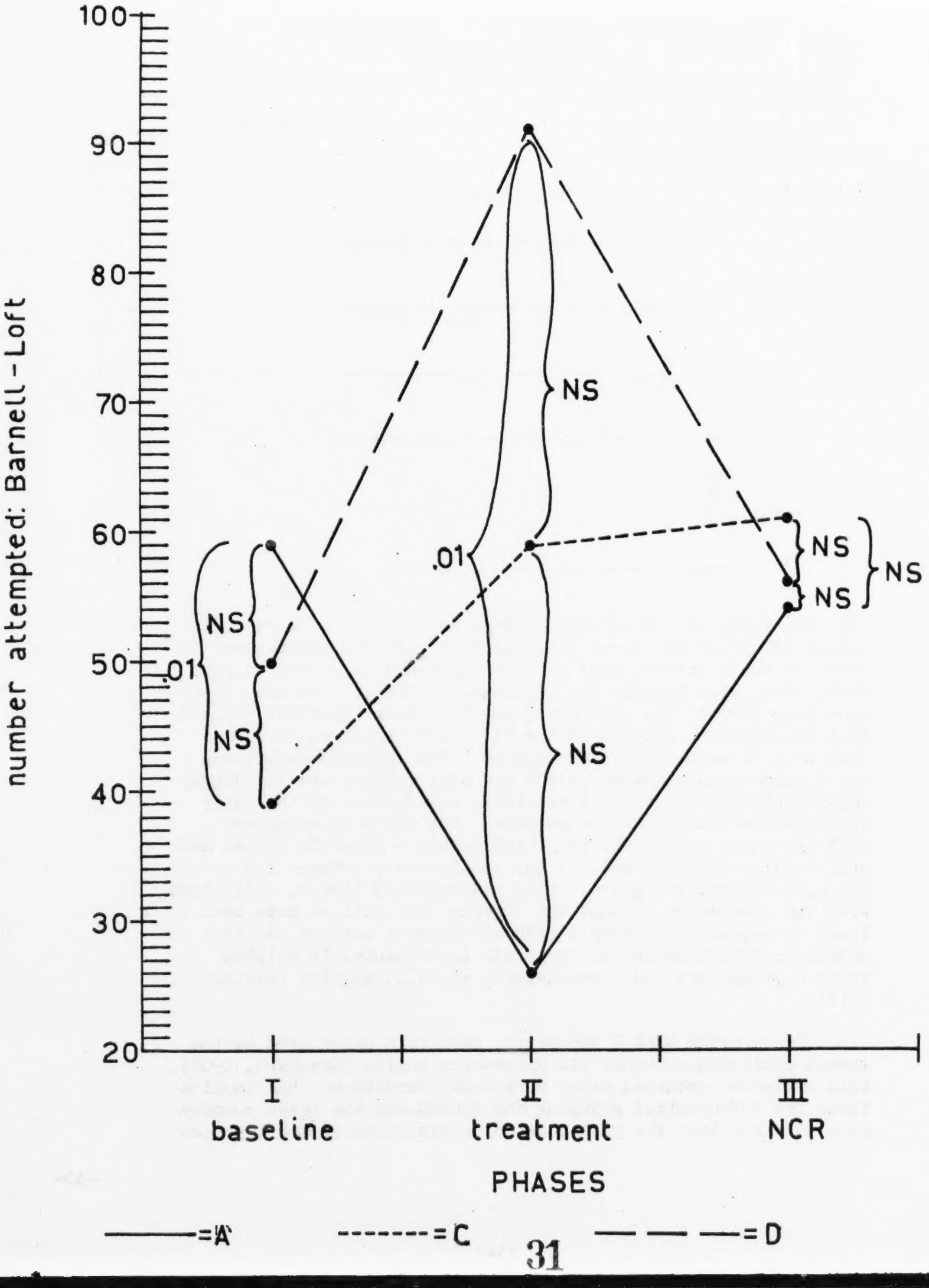

number attempted: Barnell-Loft
100
90
80
70
60
50
40
30
20
NS
.01
NS
NS
.01
NS
NS
NS
NS
NS
I
baseline
II
treatment
III
NCR
PHASES
=A
=C
=D

Table 3

Intercorrelations of Gain Scores on Spache Subtests
and
Quay Checklist Score for Each of Four Behavior Categories

Spache Tests	Quay Categories			
	C	P	I	SD
Wd Rec.	-.359*	.184	.406*	.477**
Inst. Level	.580***	-.078	.222	.377*
Ind. Level	.436**	-.015	.113	.134

* = .10
** = .05
*** = .01

The negative correlation between gain scores on Word Recognition and C and SD scores is consistent with the literature, in that one would predict that both groups would do poorly on rote tasks which Word Recognition essentially entails. On the other hand Quay (1966b) has predicted, and has shown empirical evidence that C youngsters respond much more to extrinsic reinforcement than they do to social reinforcement. The correlation between the C scores and gains on silent and oral reading are strikingly consistent with the role of extrinsic reinforcers in the daily treatment procedures of the project. The other correlations that are worth noting are the relationship between SD scores and oral reading gains as well as the relationship between I scores and word recognition gains. This relationship is also consistent with the literature in that the I group does well on rote tasks. Thus, it appears that Quay's predictions were correct and that extrinsic reinforcement was probably instrumental in helping these C youngsters and S youngsters, as well, acquire reading skills.

Considering that C youngsters have been pinpointed as the lowest achieving group of all disordered pupils (Graubard, 1968), this evidence, obtained under work-a-day conditions, has implications for differential grouping and recommends the token economy as a valuable tool for public school administrators and teachers

charged with educating conduct problem children.

Another dimension that was examined was the relationship between various data on Barnell-Loft Programmed Reading Materials and scores on the Behavior Problem Checklist. The measures used were number of frames attempted by Ss, the number correct to give an indication of accuracy, and the per cent correct to give an indication of the quality of work completed. The data were compiled during baseline, treatment, and NCR. These data are shown in Table 4.

Table 4

Intercorrelations Between Quay Checklist Category Scores
and
Number Attempted, Number Correct, and Percent Correct on
Barnell-Loft Materials for Each of Three Designated Experimental Phases.

BARNELL-LOFT VARIABLES

	C	P	I	SD
I. BASELINE				
# ATT.	-.138	-.202	.387	.063
# CORR.	-.037	-.101	-.333	-.106
% CORR.	.030	-.008	-.023	-.466*
II. TREATMENT PHASE				
# ATT.	.485**	-.116	-.068	.000
# CORR.	.520**	-.166	-.062	.067
% CORR.	-.013	-.083	-.181	.396*
III. NON CONTINGENT REINFORCEMENT				
# ATT.	.295	.164	-.312	-.510**
# CORR.	.077	.121	-.330	-.520**
% CORR.	-.422*	.046	-.192	-.079

.400 = .10*
.468 = .05**
.590 = .01***

Before treatment was instituted there was no relationship between Behavior Problem Checklist scores and academic data (with a near miss for the number of frames attempted and the I dimension) and the negative correlation between S scores and per cent of frames correct. This negative score can be interpretated as almost a deliberate attempt to do poor quality work.

It is during treatment that significant differences emerge between groups, with the C groups clearly demonstrating the highest relationship between a behavior dimension, and both numbers of frames attempted and number of frames correct. This is significant at the .05 level. Interestingly enough, just the number of attempted and number correct increase, and not the relative quality of the work as judged by accuracy of responses. In other words, under treatment, this group just did more and more work, and the number correct increases largely as a function of more done rather than increased accuracy per se. The directional difference for correlations of the SD groups of -.466 during baseline for per cent correct of frames completed and of +.396 during treatment, might also be noted as a demonstration of the relationship between treatment and the quality of work done by this group.

During NCR the significant correlation between C scores and frames attempted and number correct disappears, suggesting the possibility that treatment was causal in the increase of work rate. (A negative relationship emerges between C scores and per cent correct, suggesting that the more obstreperous children did less accurate work when direct contingencies were not in effect). A final note must be added about C scores: Those Ss with the very highest C scores did not attend school with sufficient regularity to be included in these analyses. This fact, coupled with other data presented in this report gives a certain predictive validity to the Behavior Problem Checklist, although it must be remembered that the number was small and there is not a direct correspondence between Checklist Score, and social and academic behavior. Nevertheless, this is an area that appears to be well worth pursuing.

The other striking picture appears in the SD dimension. Here, in the per cent correct portion, scores go from -.466 (Baseline) and +.396 (treatment), down to a non-significant .079 during NCR. Interestingly enough, a negative correlation -.510 and .520 emerges between the SD scores, and number of frames attempted and number of frames correct respectively, during NCR. Since per cent correct was not significantly related to the SD scores during NCR it can be assumed that number of correct declined as a function of fewer examples actually attempted. Quite possibly, these Ss, as a group, decided that as long as they were receiving reinforcement there was no sense in working for it. This is the only group that responded this way and is consistent with the activity of Socialized Delinquents as reported by Quay.

Some additional data were also gathered vis a vis the Behavior Problem Checklist and the social behaviors measured by the observers using the Becker scale. These data, presented in

Table 5, show the relationship between behavior dimensions and socially appropriate behavior.

Table 5

Intercorrelation of Quay Categories and Social Behavior During Three Experimental Phases, for Each Period.

Period 1

Phase	Quay Categories			
	C	P	I	S
Baseline	-660**	.075	.170	-.180
Treatment	-.221	-.148	.240	-.140
NCR	-.662**	-.191	.065	-.483*

Period 2

	C	P	I	S
Baseline	-.247	-.184	.118	-.040
Treatment	-.442*	-.231	.126	-.300
NCR	-.624**	-.066	.103	-.398

433 = .05*
549 = .01**

Because of the multiple baseline techniques a substantial portion of the treatment phases include traditional techniques which included withdrawal of all reinforcement. Thus, the correlations are probably minimal. During Period 1 baseline conditions, the only significant relationship between behavior dimension scores and ongoing behavior is for the C dimension, as one would expect. During treatment this relationship disappears

and the C dimension is indistinguishable from all the other dimensions. During NCR the obstreperous behavior for the C dimension reappears and the SD dimension becomes significantly related to obstreperous behavior. The coupling of the SD dimension with obstreperous behavior is parallel to that of the SD group slacking its rate of work on Barnell-Loft during NCR. For some reason, the NCR period brought out the worst in this group. The fact that Ss with C scores showed great changes during treatment is again consistent with Quay's theory that C children do not, as a rule, respond to social reinforcement, but instead need extrinsic reinforcement and novelty to motivate them.

During Period 2 the C group again shows a significant relationship with obstreperous behavior but this time during treatment at the .05 level as well as during NCR but not during Baseline. Thus there is a certain inconsistency in the data and these investigators would conclude that for the C children it is harder to maintain their attention over time and their conduct tends to deteriorate over the day.

These data suggest that groups can be differentiated on the basis of Behavior Problem Checklist scores and there is a differential response to treatment (token reinforcement) as well as to traditional classroom routines. The data also suggest that it is possible to get involved in the question of for whom is the token economy most effective, as well as the question of which classrooms need it the most. This checklist, with refinements, could develop into a powerful tool for educators in helping to group children and in providing differential methods for treating them. Thus, differences in teaching styles, and willingness and ability to work with a token economy could be paramount factors in work with C and SD children, and just not that important in work with I and P youngsters.

Discussions:

The use of operant techniques, such as token economies, seem to have a real place in the operation of schools for delinquent and pre-delinquent boys. These special schools and special classes, which go by many names, such as opportunity classes, career classes, etc., service hundreds of thousands of children throughout the nation. These classes are plagued by two major problems:

1) Management of classroom is of paramount concern, and while the per cent of deviant behavior might appear to be minimal, the quality as well as the quantity of deviant behavior is extremely difficult for teachers to handle. Violence is very much

a part of many classrooms, and many teachers leave special education because of the difficulty they have in managing classes. There is also some experimental evidence (Bruno, 1967) that those teachers who leave the classroom are those with the highest regard for individuality, nurturance, etc., and those who remain tend to be more concerned with domination, order, and authority than their counterparts who leave.

2) The amount of academic progress that delinquent children make is minimal (Lipsyte, 1970; Graubard 1964, 1968). This is probably related to teacher turnover since many teachers might leave the profession, leave working with this population, when academic gains are so minimal since pupil achievement and teacher gratification are probably highly related.

It appears then, from the results of this study, that a token economy can be an effective tool in the repertoire of teachers, since its effects were apparent in at least three of the four classes regarding social behavior, and probably in all classes regarding academic behavior. It seemed to be particularly helpful to new teachers and those who had major problems with management. The token economy seemed to lend structure, rules, and techniques which could be taught to teachers and then applied by the teachers so that management problems were reduced to a tolerable level. This is particularly important because this program took place in a public school, without using specially trained or selected teachers, and the experiment was in effect for practically an entire school year. Thus, the results of the study can probably be replicated in similar situations and these techniques do not appear to be one of the thousand auspicious ideas which cannot hold with anything but ideal conditions.

It is also important to note that the teachers were trained in a series of workshops and directly in their classrooms. It appears that the on-the-job training model is quite effective but, in the opinions of the investigators, too seldom used effectively in schools.

The token economy, on the other hand, was not a panacea and there were few days when at least one child wasn't having difficulty. The population of the school can only be characterized as volatile and while it was felt that a great deal of stability was added by the tokens and curriculum, there were numerous times when children entered the school extremely provactive and aggressive, and it did not appear as if consequences, at least for the time being, mattered to the children.

While it must be concluded that systematic use of reinforcement techniques significantly effects behavior changes in classes

for aggressive boys, it must also be stated that a large percentage of children who should have attended the school never showed up, and reinforcement techniques, like all other techniques, cannot work on children who do not show up. It is apparent that these techniques can work with children who do attend, and probably are useful for preventive work in schools so that school can be associated with positive things and truancy can be reduced. For the present, however, it is obvious that certain children cannot be enticed to attend school even with a reinforcement program, so this kind of program should not be confined to school but could be conducted in a storefront or a factory, since it appears that the school atmosphere is too powerful to counteract for some children.

What was also of great interest to the experimenters was that behavior in periods other than reinforcement periods did not deteriorate. A constant question that is asked in the field is "Won't children refuse to do work or misbehave when they aren't receiving tokens?" This study shows that while token reinforcement generally will lead to increased appropriate social behavior from baseline, it does not lead to deterioration of social behavior in periods when reinforcement is not employed. Generally, social behavior remained consistently high, although after a while when token reinforcement was not used in some cases (inexperienced teachers) it returned to baseline levels. In no case did it fall below baseline level. Thus, the available evidence indicated that reinforcement programs lead to increased appropriate behavior, and even when this specific reinforcement is withdrawn, student behavior remains at or above baseline level. The fact that behavior reverts to baseline level does not mean that the students are "not cured"; it does mean that with some children natural contingencies are not enough and "cured" e.g., improved behavior, cannot be left to chance but must be explicitly programmed.

What Kind of Reinforcement?

One of the questions this study set out to answer was which group reinforcement, individual reinforcement on groups plus individual reinforcement was most effective. Most of the data demonstrates that group delivered reinforcement was more powerful than individually delivered reinforcers. This is particularly so in Class D, and while the interpretation remains speculative, it appears that the teacher was relatively inexperienced and had a great deal of difficulty controlling the class. In this case the use of the group proved consistently superior, probably because the children could listen to the teacher under group reinforcement conditions without losing face. A previous study (Graubard, 1969b) has demonstrated the power of the group in working with delinquents

toward achieiving educational goals and this study supports the contention of Polsky (1962), Graubard (1969a), and Parsons (1954), that in working with delinquents the group must be taken into account, and Graubard's contention that programs that concentrate on individuals and do not enlist the support of the group will face almost unsurmountable odds. The use of the group appears to be the preferred technique in instituting token systems for aggressive boys.

Academic Gains

Striking academic gains did accrue to the Ss in this study. Unfortunately, the design does not permit the analysis necessary to rule out factors other than token reinforcement. What can be said is that increased academic gains did occur when Ss were presented with a combination of token reinforcement and systematic curriculum. Possibly the increased academic gains would have accrued without the token system. It can be said however, that without the token system appropriate social behavior would not have increased, but correlations between academic output and social behavior were shown to be independent for the most part. Further research is needed to clarify just how much gain can be attributed to each of these factors and an interaction between reinforcement and curriculum is necessary to achieve these gains.

Prediction

This study demonstrated that there were significant relationships between social behavior, achievement, treatment, and certain kinds of personality characteristics or traits as measured by the Behavior Problem Checklist. This opens up the possibilities of beginning large scale research on the question of prediction. For whom will token reinforcement be successful? What teachers can best use it? What is the cost effectiveness of using token reinforcement on an I child as compared to using other methods? What is the relative cost effectiveness of using token reinforcement for a P child compared to a C child? Very often teaching or treatment methods have been suggested without differentiation as to with whom they would be effective; it is now possible to look more closely at such questions.

Summary and Conclusions

A relatively recent book: Girls at Vocational High, (Meyer, Borgatta, & Jones, 1965) concluded that although it was fairly easy to diagnose the delinquency problem, effective remediation or treatment was essentially lacking.

Although this investigation did not explore out-of-school outcomes (stealing, fighting, etc.), it did effect many significant changes in behavior in school and thus demonstrated the efficacy of certain kinds of technology that can be introduced in school systems now. The token economy will not prove to be a panacea nor will it reach the very substantial number of children who are not attending school, but it can make a great difference in the lives of children and teachers now.

From this study we can conclude that:

1) Teaching is more effective with systematic reinforcement than without.
2) Treatment effects will hold up over time.
3) Social behavior does not deteriorate in periods when children do not receive reinforcement.
4) Group delivered reinforcement seems superior to individually delivered reinforcement.
5) In combination with consistent curriculum substantial reading increments can accrue.
6) Token economies seem more effective with C and S children than I and O children and further work can be done in predicting for whom token economies will be most effective.

A great deal of work needs to be done to discover how to efficiently teach these techniques to the many teachers working with a "special class," and to reach the many children who do not come to school often enough to be affected by the treatment.

Recommendation for further study -

A good number of questions remain unanswered. These include:

A) How can these behavior modification techniques be effectively taught to teachers?
B) How can we refine our prediction data to more accurate measures on effectiveness and efficiency in working with children?
C) How can we optimize the use of poor pressure to change behavior?
D) How can we reach the many children who do not attend school regularly enough to be affected by any successful program?

References

Asch, S.E. Studies of independence and conformity. Washington: American Psychological Association, 1956.

Baer, D.M., Wolf, M.M., & Risley, T. Some current dimensions of applied behavior analysis. Journal of Applied Behavior Analysis, 1968, 1, 91-97.

Becker, W.C., Madsen, C.H. Jr., Arnold, C.R. & Thomas, D.R. The contingent use of teacher attention and praise in reducing classroom behavior problems. Journal of Special Education, 1967, 1, 287-307.

Boning, R.A. Specific Skill Series, Rockville Centre, New York: Barnell-Loft Ltd., 1965.

Bruno, F. Life values: manifest needs and vocational interests as factors influencing professional career satisfaction among teachers of emotionally disturbed children. (Unpublished doctoral dissertation, Wayne State University), 1968.

Bushell, D., Wrobel, P.A., & Michaelis, M.L. Applying "group" contingencies to the classroom study behavior of preschool children. Journal of Applied Behavior Analysis, 1968, 1; 55-63.

Clark, M., Lachowicz, J., & Wolf, M.M. A pilot basic education program for school dropouts incorporating a token reinforcement system. Behavior Research and Therapy, 1968, 6, 183-188.

Cloward, R.A., & Ohlin, L.E. Delinquency and Opportunity. Glencoe, Ill.: The Free Press, 1960.

Conant, J. B. Slums and Suburbs. New York: McGraw-Hill, Inc., 1961.

Dunn, L. Peabody Picture Vocabulary Test. Minneapolis: Minnesota Test Bureau, 1959.

Eiduson, Bernice, Johnson, Trinidad, and Rottenberg, D. Comparative studies of warning problems seen in five child clinics. American Journal of Orthopsychiatry, 1966, 36, 829-839.

Graubard, P.S. "The Extent of Academic Retardation in a Residential Treatment Center," Journal of Educational Research, 1964, 58, 78-80.

References (Cont'd)

Graubard, P.S. The use of the Peabody Picture Vocabulary Test in the prediction and assessment of reading disability in disturbed children. Journal of Educational Research, 1967, 61, 3-5.

Graubard, P.S. "An Investigation of Reading Correlates of Emotionally Disturbed and Socially Maladjusted Children: The Relevance of a Classification Schema to Educational Characteristics." A report submitted to the U.S.Office of Education, 1968.

Graubard, P.S. Children against schools. (ed) Chicago: Follett Educational Corp., 1969a.

Graubard, P.S. Utilizing the Group in Teaching Disturbed Delinquents to Learn. Exceptional Children. 1969b, 267-272.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953, 292-297.

Lipsyte, S. Director, Bureau of Education for Maladjusted Children, New York City Board of Education, personal communication, 1970.

Meyer, H.J., Borgatta, E.F., & Jones, W.C. Girls at Vocational High. New York: Russell Sage Foundation, 1965.

O'Leary, K.D., Drabman, R. Token Reinforcement Programs in the Classroom: A Review. State University of New York at Stony Brook: in press.

O'Leary, K.D., Becker, W.C., Evans, M.B., & Saudargas, R.A. A token reinforcement program in a public school: A replication and systematic analysis. Journal of Applied Behavior Analysis, 1969, 2, 3-13.

Parker, D.H., & Scannell, G. SRA Reading Laboratory, Chicago: SRA, 1963.

Parsons, T. Essays in Sociological Theory, (revised ed.) Glencoe, Ill.: The Free Press, 1954.

Peterson, D.R. "Behavior Problems of Middle Childhood." Journal of Consulting Psychology, 1961, 25, 205-209.

References (Cont'd)

Phillips, E.L. Achievement place: token reinforcement procedures in a home-style rehabilitation setting for "pre-delinquent" boys. Journal of Applied Behavior Analysis, 1968, 1, 213-224.

Polsky, H.W. Cottage Six - The Social System of Delinquent Boys in Residential Treatment. New York: Russell Sage Foundation, 1962.

Quay, H.C. Personality patterns in pre-adolescent delinquent boys. Educational and Psychological Measurement. 1966a, 26, 99-210.

Quay, H.C. Dimensions of problem behavior in children and their interaction with approaches to behavior modification. Kansas Studies, 1966b.

Quay, H.C. Personal communication. 1966c.

Quay, H.C., Morse, W.C., and Cutler, R. Personality patterns of pupils in special classes for the emotionally disturbed. Exceptional Children, 1966a, 32, 297-301.

Spache, G.D. Diagnostic Reading Scales. Monterey, California: California Test Bureau, 1963.

Staats, A.W., & Butterfield, W.H. "Treatment of Nonreading in a Culturally Deprived Juvenile Delinquent: Application of Reinforcement Principles," Child Development, 1965, 36, 925-942.

Wechsler Intelligence Scale for Children. New York: Psychological Corporation, 1949.

Wolf, M.M., Giles, D.K., & Hall, V.R. Experiments with token reinforcement in a remedial classroom. Behavior Research and Therapy, 1968, 6, 51-64.

APPENDIX - A - Quay Behavior Problem Checklist

Col. No.	Please complete each question carefully.
(1-8)	1. Name (or Number) of child _____
(9-10)	2. Age (in years) _____
(11)	3. Sex _____(M 1, F 2)
(12)	4. Father's Occupation _____
(13)	5. Name of person completing this checklist _____
(14)	6. What is your relationship to this child? (circle one) a. Mother b. Father c. Teacher d. other _____ (Specify)
(15-16)	7. School _____
(17)	8. Grade _____

Please indicate which of the following constitute problems, as far as this child is concerned. If an item does not constitute a problem, encircle the zero; if an item constitutes a mild problem, encircle the one; if an item constitutes a severe problem, encircle the two. Please complete every item.

(18)	0	1	2	1. Oddness, bizarre behavior
(19)	0	1	2	2. Restlessness, inability to sit still
(20)	0	1	2	3. Attention-seeking, "show-off" behavior
(21)	0	1	2	4. Stays out late at night
(22)	0	1	2	5. Doesn t know how to have fun; behaves like a little adult
(23)	0	1	2	6. Self-consciousness; easily embarrassed
(24)	0	1	2	7. Fixed expression, lack of emotional reactivity
(25)	0	1	2	8. Disruptiveness; tendency to annoy & bother others
(26)	0	1	2	9. Feelings of inferiority
(27)	0	1	2	10. Steals in company with others
(28)	0	1	2	11. Boisterousness, rowdiness
(29)	0	1	2	12. Crying over minor annoyances and hurts
(30)	0	1	2	13. Preoccupation; "in a world of his own"
(31)	0	1	2	14. Shyness, bashfulness
(32)	0	1	2	15. Social Withdrawal, preference for solitary activities
(33)	0	1	2	16. Dislike for school
(34)	0	1	2	17. Jealousy over attention paid other children
(35)	0	1	2	18. Belongs to a gang
(36)	0	1	2	19. Repetitive speech
(37)	0	1	2	20. Short attention span
(38)	0	1	2	21. Lack of self-confidence
(39)	0	1	2	22. Inattentiveness to what others say
(40)	0	1	2	23. Easily flustered and confused
(41)	0	1	2	24. Incoherent speech
(42)	0	1	2	25. Fighting
(43)	0	1	2	26. Loyal to delinquent friends
(44)	0	1	2	27. Temper tantrums
(45)	0	1	2	28. Reticence, secretiveness
(46)	0	1	2	29. Truancy from school
(47)	0	1	2	30. Hypersensitivity; feelings easily hurt
(48)	0	1	2	31. Laziness in school and in performance of other tasks
(49)	0	1	2	32. Anxiety, chronic general fearfulness
(50)	0	1	2	33. Irresponsibility, undependability
(51)	0	1	2	34. Excessive daydreaming
(52)	0	1	2	35. Masturbation
(53)	0	1	2	36. Has bad companions
(54)	0	1	2	37. Tension, inability to relax
(55)	0	1	2	38. Disobedience, difficulty in disciplinary control
(56)	0	1	2	39. Depression, chronic sadness
(57)	0	1	2	40. Uncooperativeness in group situations
(58)	0	1	2	41. Aloofness, social reserve
(59)	0	1	2	42. Passivity, suggestibility; easily led by others
(60)	0	1	2	43. Clumsiness, awkwardness, poor muscular coordination
(61)	0	1	2	44. Hyperactivity; " always on the go"
(62)	0	1	2	45. Distractibility
(63)	0	1	2	46. Destructiveness in regard to his own &/or other's property
(64)	0	1	2	47. Negativism, tendency to do the opposite of what is requested
(65)	0	1	2	48. Impertinence, sauciness
(66)	0	1	2	49. Sluggishness, lethargy
(67)	0	1	2	50. Drowsiness
(68)	0	1	2	51. Profane language, swearing, cursing
(69)	0	1	2	52. Nervousness, jitteriness, jumpiness; easily startled
(70)	0	1	2	53. Irritability; hot-tempered, easily aroused to anger
(71)	0	1	2	54. Enuresis, bed-wetting.
(72)	0	1	2	55. Often has physical complainss, e.g. headaches, stomach ache

APPENDIX - B - Beck CODING CATEGORIES with MODIFICATIONS

Symbols	Class Label	Class Definitions
A.	Behaviors Incompatible with Learning: General Categories	
X	Gross Motor Behaviors	Getting out of seat; standing up; running; skipping; jumping; walking around; rocking in chair; disruptive movement without noise; moving chair to neighbor.
X-AB-	Out of Room	
N	Disruptive noise with objects	Tapping pencil or other objects; clapping; tapping feet; rattling or tearing paper. (Be conservative, only rate if could hear noise with eyes closed. Do not include accidental dropping of objects or noise made while performing X above.)
A	Disturbing others directly and aggression	Grabbing objects or work; knocking neighbor's book off desk; destroying another's property; hitting; kicking; shoving; pinching; slapping; striking with object; throwing object at another person; poking with object; attempting to strike. Bantering.
-AF-	Fighting	
L	Looking	Turning head or head and body to look at another person; showing objects to another person; attending to another child. (Must be of 4 seconds duration to be rated. Not rated unless seated.)
B	Blurting out, Commenting, and Vocal noise	Answering teacher without raising hand or without being called on; making comments or calling out remarks when no question has been asked; calling teacher's name to get her attention; crying; screaming; singing; whistling; laughing loudly; coughing deliberately loudly. (Must be unitrected to another particular child, but may be directed to teacher.)
T	Talking	Carrying on conversations with other children when it is not permitted. (Must be directed to a particular child or children.)

-44-

APPENDIX -B - Becker CODING CATEGORIES with MODIFICATIONS (Cont'd)

Symbols	Class Label	Class Definitions
O	Other	Ignoring teacher's question or command; doing something different from that directed to do (includes minor motor behavior such as playing with pencil when supposed to be writing.) (To be rated only when other ratings not appropriate.) Day Dreaming, Mapping, Sleeping, Stripping, Undressing above the waist.
Ab	Absent from school	Out of room
EA TEA	Excused absence	Must be out of the classroom (child-initiated) as monitor, to bathroom, etc.
E	Expelled	Sent to office, guidance counselor, etc., punitively; or other punitive arrangements.
SX		Masturbation; Feeling someone else.
S - TEA		Child is doing an activity different from the others-- But with Teacher's approval, Ex. Child has headache-- "lay your head on desk," or "color", or "Read Comic Book"

B. Special Categories

Idiosyncratic behavior

_____

_____

_____

_____

_____

_____

_____

_____

_____

APPENDIX - B - Becker CODING CATEGORIES with MODIFICATIONS(Cont'd)

Symbols	Class Label	Class Definitions
C.	Relevant Behavior	
S.	Relevant Behavior	Studying, writing, eyes on task, answering questions, listening, raising hand, following teacher's directions. (Must include whole 20 seconds except for orienting responses of less than 4 seconds.)

Observers: Tape stop watch to clipboard. Start watches together and check for synchronization every 10 minutes. Observe each child for 20 seconds and take ten seconds to record the classes of behavior which occurred during the 20 second period. Wait ten seconds and observe next child.

APPENDIX -C- Some Extra-Experimental Considerations
for Experimenters in the Public Schools

The special services school which we entered in September 1968 had hosted experiments and special projects for most of its five years, and thus appeared to be receptive to innovation in approaches with its students. In our first training session with the teachers working with us, most of them expressed an interest in participating in and learning "something new." It is no more to impugn the sincerity of this receptivity than it is to impugn the sincerity of the investigators in undertaking the study to say that a common goal of "wanting to help children" leaves ample room for disparate agendas of how to help them. The school was not only open to special projects but, having known them most of its life and being desperately under-equipped in materials, heavily relied on them. Without the programmed materials we brought, which greatly eased our entrance and functioning in the school, most of the classrooms in which we worked would have been almost bare. The concrete value of needed curriculum was a vital and immediate benefit of the study to the teachers, when the service offshoot of the experiment, and the long-term payoff of testing a new method, were obscure.

The inexperienced teachers, in a school where supervisory time was at a premium, singled out the training and conference time with the study's staff as a greatly needed benefit which encouraged fuller application of their resources in the experiment.

Being interested in learning something new is not a guarantee of being interested in doing something new day in and day out. The same statement from a highly successful, relaxed teacher and from an apprehensive novice teacher can mean very different things: the first year teacher may have the incentive born of urgency and even desperation, but the disadvantage of being so harried that she does not have time to explore the technique and use it creatively. The skillful, experienced teacher has the advantage of being likely to succeed with virtually any technique she tries, but may have the disadvantage of an understandably large investment in the techniques to which she is accustomed. Particularly in a school such as this, success is hard to come by and the "John Henry" syndrome is a likely result of achieving it. (The "John Henry" syndrome is defined by these investigators as a pride in having forged one's way oneself without the aid of new-fangled methods or outside advice.)

If she has been thrown on her own resources and devised her own solutions, the experienced teacher may find anything that is not "doing it herself" and object of suspicion and resistance. Thus, her class's real progress in an unfamiliar, experimenter-designed framework, may be

-49-

difficult for her to recognize. Conducting an experiment and armed with scientific principles and a rigorous attitude towards their application, we asked for de-emphasis of "personality" teaching; but personality teaching may be one of the teacher's most important sources of satisfaction. We cannot offer a ready resolution of this problem where experimental considerations limit technique variations, beyond suggesting that teachers who have less of a stake in their own methods, with more in concrete improvement of their teaching, are probably more willing and able participants in a study.

We recommend the utmost clarity on what a study aims for, specifically detailing what it will require of and give participants, and what they require of and can give to it. Research designs such as variations of the multiple-baseline used here which could allow potential participants to try out the techniques before their or the experimenter's commitment might avert strain later on. Designs where changes could be timed with changes in the students would also be helpful. More sophisticated technology than was at our disposal should be provided to make sure the teacher's work load -- e.g., record-keeping -- is not increased out of proportion to her gains from using the new techniques. Such an apparently simple issue as record-keeping created possibly the most difficult barrier between experimenters and teachers, until streamlined techniques and extra help were arranged.

In a school where tension and pressure was already high (as detailed below) there is a tendency for the rest of the faculty to see the participating teachers both as a specially privileged elite and as breakers-of-rank. We would have done well to pay more attention to this aspect, explaining the experiment more fully to the whole school, seeing beyond the research and making, at some point in the year, our consultation available to some extent to any teacher who wished it, or perhaps arranging consultation for other staff members with one of our more expert trainees.

Though it does not figure directly in the statistics of this report, the circumstances of 1968-69 in the schools of New York City had a great impact on our daily functioning. This was the year of the school strike. It is a credit to everyone concerned that the study operated at all. In the school in which we worked, all but two of the teachers worked throughout the strike. Thus there was not the split down the middle present in many faculties when the strike ended, but there was, with such an inauguration, an atmosphere of heightened tension in the school throughout the year. Racial animosity was surprisingly rare, but it surrounded the school, in the newspapers, several blocks

uptown, in the air. Any day might mean a new strike or a new explosion somewhere, and some of our teachers seemed to be struggling with just how militant they should be; we all wondered from time to time if this was where our energies should be going. The former principal of this school spearheaded the drive for community control in Brooklyn. Many children, already likely truants, got into the habit of not attending school over the six-week strike, and never abandoned it. The school, drawing from throughout Manhattan but situated in a middle-class community, was under pressure from neighborhood merchants who complained of shoplifting, and residents, who complained of harrassment from the imported children. A custodians' strike left the heating system in-operative for a long cold stretch, and wearing hats and coats was not conducive to learning. The wide range of teaching abilities and orientations, from the skilled and sympathetic teacher who chose this school as a meaningful challenge, to the barely competent brutal disciplinarian who chose this school because it needed teachers, impeded intra-faculty cooperation and gave one the feeling of moving from the 20th Century to the Dark Ages in a walk down the hall.

Overeagerness in our first experience with an experiment in a public school may have prompted us to tolerate the ambivalence of two teachers, who eventually dropped out of the project, much too long. Although these teachers never really participated in the study, their wavering -- in one case, for months -- resulted in our investing much time and effort trying to devise programs specifically tailored to the ambivalence. Had we accepted earlier that strong ambivalence is probably insuperable in this kind of situation, we might have spent that time much more fruitfully.

Also, despite the reinforcement techniques we espoused, often our work with the teachers ran counter to them. When a teacher was carrying out the reinforcement techniques smoothly, she received less of our attention and time than when she ran into trouble. Of course, the teacher who is having a hard time probably requires more observation, modelling, feedback, and conference time than the teacher who is not, particularly if research goals are in the front of one's mind. However, attention can and should be given to the succeeding teacher. In our situation, she might have given (preferably with pay) workshops for other school personnel in the use of reinforcement techniques, with her own class as a demonstration. With frequent outside observers inevitable, she could have met with them to explain the program from her point of view. Probably such a teacher could participate in some of the investigators' consultations to other schools, and if her investment and abilities allowed, co-author a paper on some specific aspect of the experiment in which she participated.

In this year which certainly demanded great sensitivity from everyone involved in education in New York City, we hired a sensitivity trainer. As it turned out, this aspect of our work, if fashionable, was both superfluous and in some ways detrimental. The teachers had enlisted to learn certain techniques and participate in an experiment, and not to undergo sensitivity training. Time spent in that training, which resulted in frequent confusion, would better have been spent in more task groups. This is not to say that such training cannot be valuable, but as an adjunt to a scientific study, it is questionable. With a lack of clarity as to what it was for and how this fit into the experiment, the trainer became a discriminative stimulus for complaints. His self-styled role as a funnel for communication, particularly communication on problems, may have served short-term gains but also may have elicited more problems than were really present, and deflected the natural flow of communication on these and other issues. Certain aspects of this training, which we feel would be possible and perhaps more natural without specific sensitivity training, were helpful. For example, the five minute period in an early training session of separating into groups of two for a meeting, not based on task or credentials, was an excellent ice-breaker. The investigators benefitted from some sessions spent on the question of their functioning productively as a team.

Relating genuinely with school personnel on matters other than the study was instrumental in teachers seeing experimenters as more than "scientists" and experimenters seeing teachers as more than agents of treatment techniques. The personal friendships which continued after-hours were also important. The project office, with its full coffee pot and available telephone (getting to the school telephone is often a rare achievement) was always open to the participating teachers. Kuypers, Becker, and O'Leary pointed out in "How to Make a Token System Fail" (Exceptional Children, 1968), that the role of the data collector has other aspects as crucial as reliable observation. There, noisy gum-chewing by observers was instrumental in the failure mentioned in the title. It was apparent that the three ladies who collected data in our study were an invaluable asset to the experiment. It is difficult to specify what seems intangible and unreplicable, i.e. "personality,"but certain factors can be isolated. The "indigenous paraprofessionals," clearly identifying with the study, also enjoyed a special rapport with the participating teachers. For the paraprofessionals and most of the teachers, the "inner-city" was more than just an area in which they had chosen to apply their skills. In most instances, the differences between them lay in that the teachers had had the good fortune to attend college, and put that education to use, and the data collectors had not. Although the paraprofessionals had a clear grasp of what the study was about, they

were not given to scientific terminology and, like the teachers, were certainly more people than experiment-oriented. Although university personnel who would venture into a public school setting cannot be solely experiment-oriented, because of their task, they are bound to carry some of the ivory tower with them.

The observers in our study, besides being extremely competent at collecting and compiling data, and helping teachers and the study with record-keeping, were also a very crucial bridge which promoted understanding in all concerned. To have used graduate students, who might have initially appeared to require less training, would have been to lose this bridge; they, at that critical stage of their lives, probably would have brought, even more than the experimenters, the ivory towers and the scientific jargon. Time and time again, the warmth, humor, and refreshing good sense of our data collectors cut through the school's experiment barriers.

In addition to the data collectors, who, more than anyone else, had to interact every day with teachers and pupils, two other factors contributed, we believe, to the success of this experiment. Administrative involvement in the project was instrumental in holding the operation together in difficult periods. In exchange we tried to be as helpful as we could in every area of our competency. Meetings were regularly scheduled with the school principal and our project kept him informed of every step of our operation. Sometimes the meetings would last for only a few minutes, but they were held consistently, and problems were not allowed to grow. The second factor which we considered especially helpful was that whenever and wherever school action and success of personnel could be publicized for project-related work, they were; the school's cooperating teachers' efforts were invariably redoubled following such public recognition.